Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US05/000201

International filing date: 06 January 2005 (06.01.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/534,520
Filing date: 06 January 2004 (06.01.2004)

Date of receipt at the International Bureau: 03 March 2005 (03.03.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

1286149

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

February 16, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: ***60/534,520***
FILING DATE: ***January 06, 2004***
RELATED PCT APPLICATION NUMBER: ***PCT/US05/00201***

Certified by

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

PTO/SB/16 (08-03)
Approved for use through 07/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No.

INVENTOR(S)		
Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Justin	picard	Providence, RI

Additional inventors are being named on the ____ *separately numbered sheets attached hereto*

TITLE OF THE INVENTION (500 characters max)

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

☐ Customer Number:

OR

☑ Firm or Individual Name	Jian Zhao, MediaSec Technologies				
Address	10 weybosset st., Ste 501				
Address					
City	Providence	State	RI	Zip	02903
Country		Telephone	401-2723388	Fax	401-272-4884

ENCLOSED APPLICATION PARTS *(check all that apply)*

☑ Specification *Number of Pages* 21 ☐ CD(s), Number ____

☐ Drawing(s) *Number of Sheets* ____ ☐ Other (specify) ____

☐ Application Date Sheet. See 37 CFR 1.76

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

☑ Applicant claims small entity status. See 37 CFR 1.27.

☑ A check or money order is enclosed to cover the filing fees.

☐ The Director is herby authorized to charge filing fees or credit any overpayment to Deposit Account Number: ____

☐ Payment by credit card. Form PTO-2038 is attached.

FILING FEE Amount ($)

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☑ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are: ____

Respectfully submitted,

SIGNATURE ____ Date 1/6/2004

TYPED or PRINTED NAME Jian Zhao REGISTRATION NO. ____ *(if appropriate)*

TELEPHONE 401-272 3388 x22 Docket Number: ____

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PROVISIONAL APPLICATION COVER SHEET
Additional Page

PTO/SB/16 (08-03)
Approved for use through 07/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Docket Number

INVENTOR(S)/APPLICANT(S)		
Given Name (first and middle [if any])	Family or Surname	Residence (City and either State or Foreign Country)
Jian	Zhao	Rumford, RI

Number 1 of 1

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

medias01.014

Provisional Patent Application

Patent Application for

Visible authentication patterns for printed documents

Extension I

Inventors: Justin Picard, Dr. Jian Zhao

Assignee: MediaSec Technologies GmbH

Visible authentication patterns for printed documents

Cross references to related applications

The present application is an extension and improvement of PCT/US03/15188, Visible authentication patterns for printed documents, filed 14 May 2003.

Extensions and Improvements of Algorithms:

1. Histogram conversion as a core part of the algorithm

The effect of printing and scanning can be seen as corresponding essentially to adding a certain amount of random noise to each pixel value. As a copy is always evaluated after two printing-scanning of the document (and the CDP), it is expected that more noise will be added for a copy than for an original. A possible way to discriminate copies from originals would then be to measure a simple difference between the digital original and the test pattern in the spatial domain: a higher distance would be expected if the test pattern is a copy. That is, if, for the pixel located at column I and row j, x(i,j) an y(i,j) are respectively the pixel value for the digital original and the test pattern, a measure of distance between the digital original and test pattern is:

$$D=\mathrm{Sum\_j\ Sum\_i}\ |x(I,j)-y(I,j)|^{p} / (N*M) \quad \text{(eq. 1)}$$

Where p is an arbitrary positive number, and N and M are the width and height on the patterns in pixels. As said before, It is expected that the distance D will always be higher for an original than for a copy. However, though it can be proven mathematically that this measure is nearly optimal for discriminating copies from originals, it is not applicable without some processing of the scanned, restored pattern. The reason is that through printing and scanning, patterns undergo luminance shift and in general non-linear transformation of the pixel values; these transformations vary upon the printer and scanner, and even on the printing/scanning parameters. Generally, the spectrum of the CDP pixel values will be compressed by printing-scanning, the extreme values of the spectrum being much more rare than in the original digital CDP.

It is however possible to revert that transformation of the CDP spectrum by using a a so-called "histogram conversion function", ie a function that modifies each individual pixel of a scanned restored CDP, such that its histogram is equivalent to the histogram of an original digital CDP. This transformation function is generally estimated on a set of printed CDPs in the calibration step. The transformation function is generally fixed, and is applied on each test CDP after it has been restored. An example histogram conversion function is shown in Figure 1, and Figure 2 shows the effect of an histogram conversion on a scanned restored CDP.

The average luminance of a printed CDP has typically some variation, due to different lighting conditions in the scanner and/or different amount of ink injected in the paper. To minimize the effect of this variability which occurs naturally and is not possible to control, it is possible to add/subtract a fixed value to each pixel in the CDP, such that its average luminance comes to a fixed value, eg 127. For example, if the average value of pixel luminance is 118, then 9 is added to each pixel. This shift adjustment of pixel luminance is typically applied before the histogram conversion.

After this transformation is applied, the pixel values of the scanned-restored CDP will have the same spectrum as the pixel values of the digital CDP. There are therefore comparable, and equation 1 can be applied. if f() is the histogram conversion function, the distance D is given by:

D=Sum_j Sum_i |x(I,j)-f(y(I,j))|^p / (N*M) (eq. 2)

Note: this distance function is just one example; sevarl other ditsnace functions might be used.

ADDITION and Variations:

For certain applications, certain variations can occur in the printing-scanning environment. This variations can typically occur in:

- different printers used to produce the document containing the CDP
- different papers used to print the CDP
- different scanners used to scan the CDP

One such application would be a postage meter (!) where CDPs can be printed on different types of envelope (with paper of different properties) and sometimes from different printers.

For such applications, the histogram conversion function could be different for each printer-paper-scanner combination. Applying the wrong histogram conversion function would result generally in an overestimation of the distance between the test and digital CDP. One solution would then be to use several histogram conversion function, and when verifying a CDP, try each of the histogram conversion function and consider only the CDP with the lowest distance to the digital CDP. If other parameters are dependent on the specific printer-paper-scanner combination (eg the threshold), then the selection of the "best" histogram conversion function would imply the selection of the other parameters.

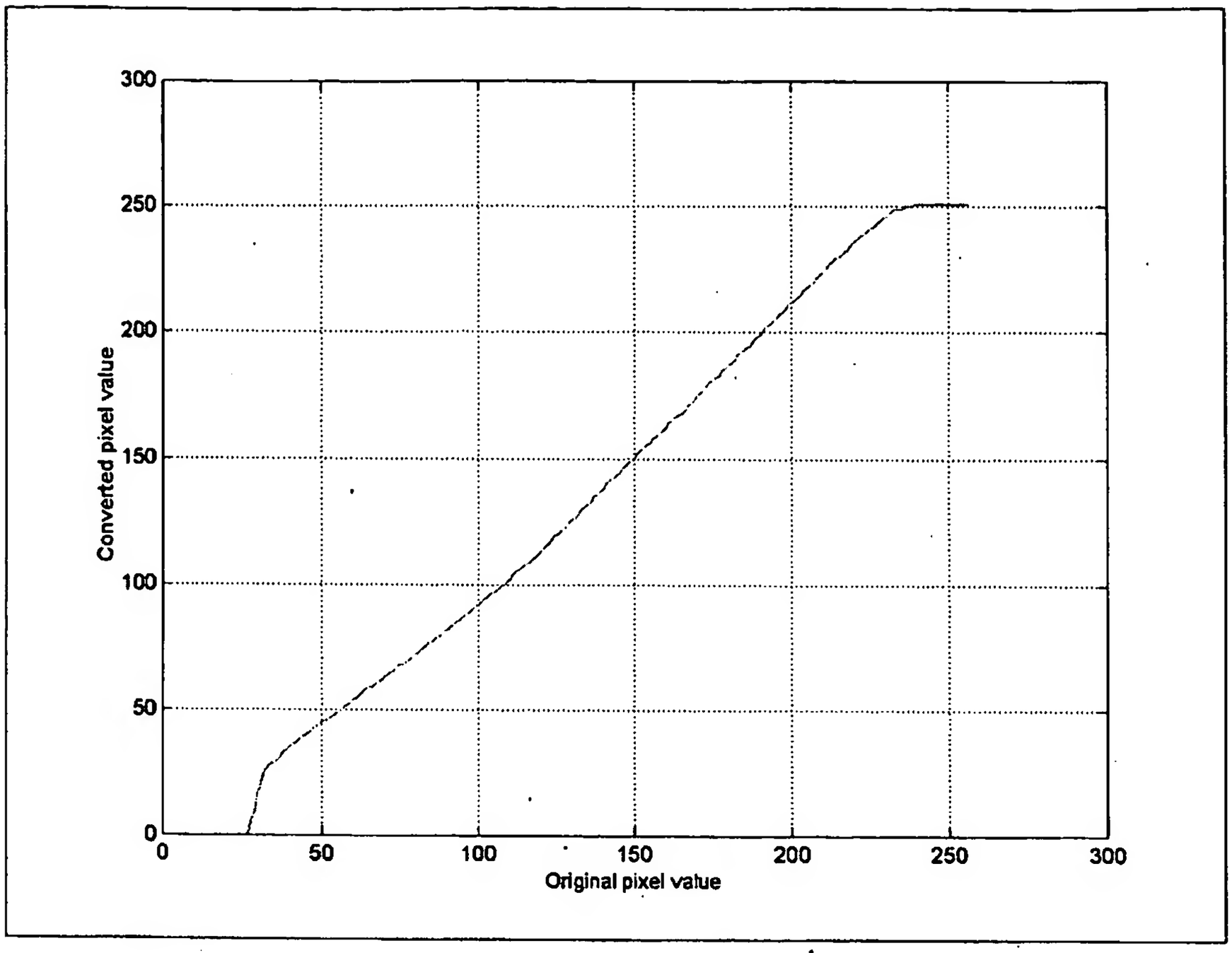
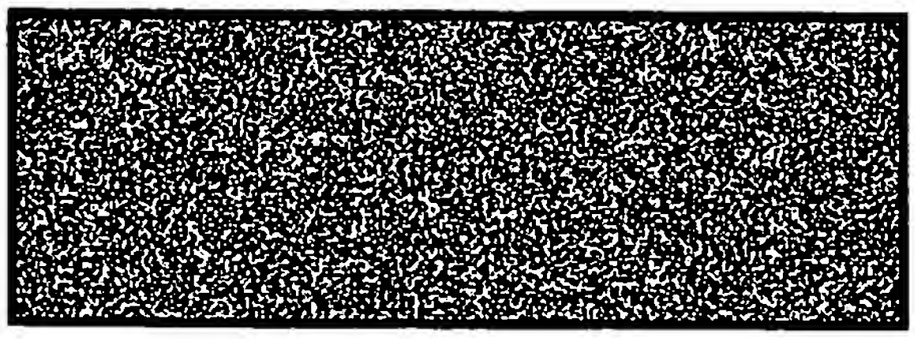
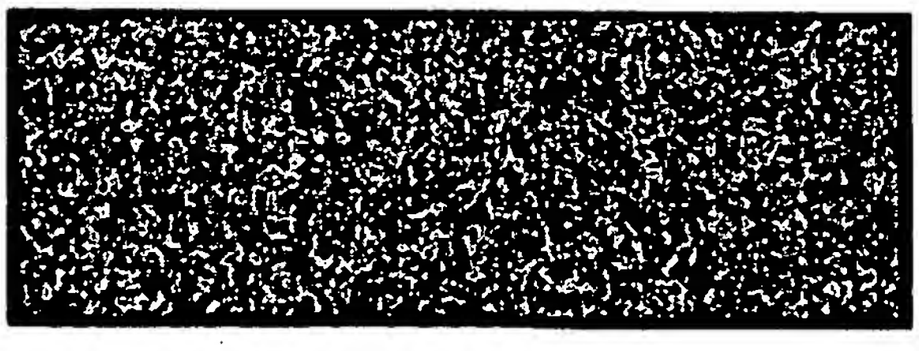


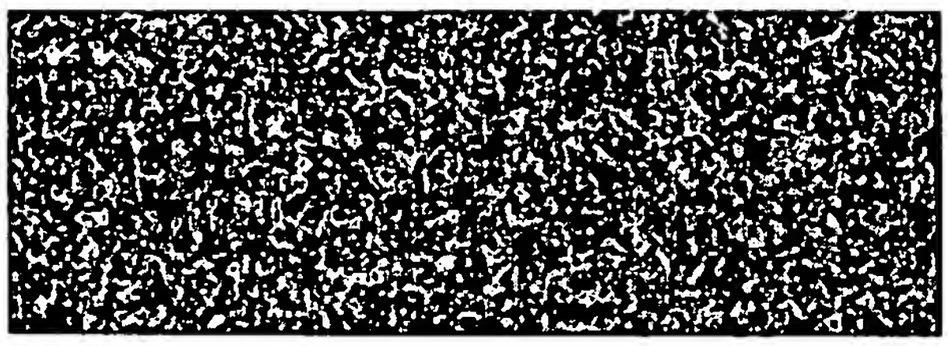
Figure 1: Histogram conversion function



(a)



(b)



(c)

Figure 2: (a) original digital CDP, (b) this original digital CDP after printing, scanning and restoration, and (3) the restored CDP after histogram conversion.

2. New method to insert information in CDP relying on CDP properties

The general scheme to insert and later on detect information in a CDP is as follow:

At creation time:

1) create original digital CDP

2) apply modifications to the original digital CDP in order to insert message M

At detection time:

a) read message M from scanned CDP

b) create original digital CDP by following steps 1) and step 2) above

c) compare the scanned CDP with original digital CDP.

The message M that is inserted in the CDP may initially be in any arbitrary form, for example: a number, some text, a date, or any combination of these. In all cases, the message can be converted in an equivalent binary representation, ie a binary sequence of "0" and "1". An error-correction code is typically applied to this sequence, which can then be repeated several times depending on the available space. A key-based cryptographic algorithm is typically applied to pseudo-randomly change the position and value of each bit; knowning the key, the transformation is revertible.

The message M can be inserted in a variety of ways. Contrarily to a digital watermark, there is no visibility constraint to respect because a CDP appears merely like a meaningless noisy graphic to the naked eye. However, in modifying the pixel values of the CDP, it is desirable to (1) avoid corrupting the CDP properties, eg keep the entropy to a maximum value, and to (2) be able to revert the modifications to the CDP after reading the information. Indeed, only if the message is read can the original digital CDP be re-generated for comparison with the scanned CDP.

There is one method to insert information in a CDP that fully preserves the histogram property of the CDP, at least for the case where all pixel values are

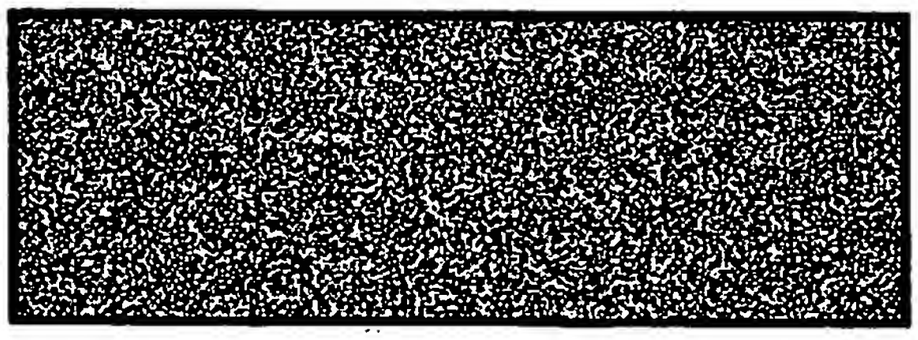
equi-probable. The principle is to divide the CDP in blocks of fixed size, e.g. 4X4, and for each block to embed a zero, each of the pixel value in the block is left unchanged, while to embed a one, each of the pixel value is inverted, i.e. pixel value x becomes x'=255-x. For example, to embed a '0' in a block with the following values:

243	228	210	236
59	195	114	189
155	117	158	45
124	5	203	104

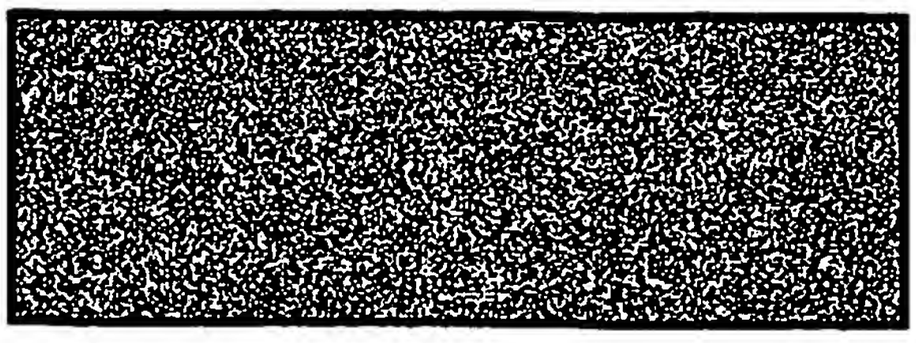
the pixel values become:

12	27	45	19
196	60	141	66
100	138	97	210
131	250	52	151

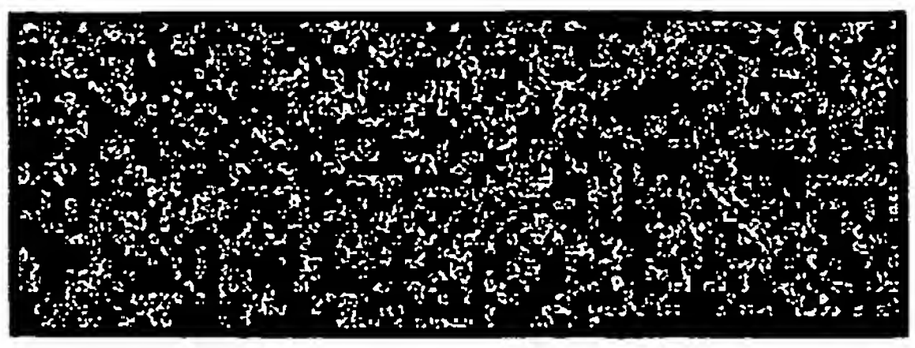
It can be easily verified that the histogram or frequency distribution of pixel values of a CDP with inserted information remains unchanged. Figure 3 shows a CDP generated with the key "test", a CDP generated with the same key with added information (integer value 123456789), and a difference image. The black areas (actually blocks of 4x4 pixels) of the difference image corresponds to the pixel values which have not been modified (inverted) by the insertion of information: they correspond to blocks where a '0' is embedded. Obviously, the noisy like, modified areas correspond to area where a '1' is embedded



(a)



(b)



(c)

Figure 3: (a) CDP without information, (b) CDP generated with the same key and information "123456789", and (c) difference image between (a) and (b)

At detection time, in a first step the original digital CDP without information is re-generated using the secret key. Then each block of the scanned, restored and histogram converted CDP is compared with (1) the corresponding block of the original digital CDP, and (2) the same block with inverted pixel values. Different comparison functions can be used: the Euclidean distance, the absolute distance, etc. Then if the closest block is the inverted one, then the bit value of the block is assumed to be '1', and '0' otherwise.

For example, suppose that for the block with pixel values above, the block of the scanned-restored-converted CDP are:

44	36	24	10
198	20	167	83
97	159	135	198
106	299	10	172

Then the absolute difference would be:

For a '0': (|243-44|+ |228-26|+...+|104-172|)/16=132.81

For a '1': (|12-44|+|27-26|+...|151-172|)/16=22.93

In that, case the detected bit would obviously be a '1'.

If a bit is embedded several times at different places in the CDP, it is possible to keep track of the distances for a '0' and a '1' for each location, such that the contribution of each block to the final decision on the bit value is weighted. This way, a block where a distance of 55.32 is found for a '0' and 51.34 for a '1' would contribute less than the bock whose computations are shown above, where the evidence in favor of a '1' is much stronger (distance of 22.93 vs 132.81).

3. Method of detecting the position of a CDP on a document with no a priori knowledge n its location

It is not always possible to scan precisely the area where the CDP is in a document. It can be either because the application has to support documents with different formats and/or a CDP placed at different location on it; or because the end user placing the document on the scanner does not know where to place document on it; or simply because there is a natural variability in each scan, and patterns in the document close to the CDP interfere with the CDP. In the "worst case application", the full area of a letter-size scanner is scanned, and the CDP can potentially be located anywhere and with any orientation on the scanner.

It is however possible to take advantage of the general statistical properties of the CDP to locate it. The property that distinguishes the CDP from most other image or document features is the spreading of its histogram. In general, each of the 256 pixel values of original digital CDP is equally probable. Though the printing and scanning of the CDP does modify the distribution of its pixel values, it is nevertheless highly specific. By printing and scanning a certain number of CDPs (in the so-called calibration process), it is possible to make an estimate of the average distribution of pixel values. Figure 4 below shows one such distribution, which is called the "template histogram".

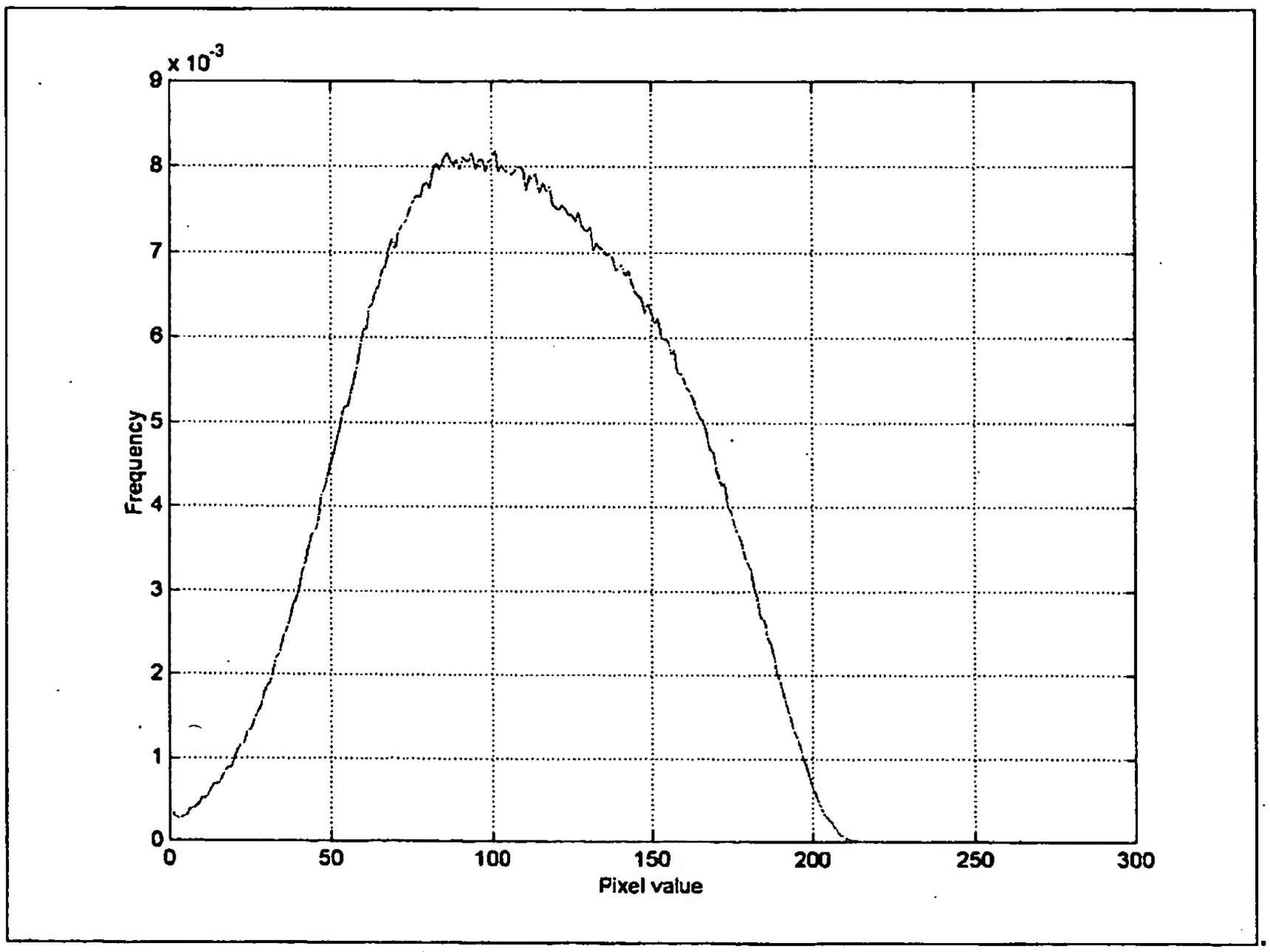


Figure 4: example frequency distribution of pixel values

The image of a scan of a document can be divided in blocks, typically of size 50x50 pixels. For a letter-size scanner scanning at 300dpi, the scanned image of 3300x2550 pixels is divided into 66x51=3366 blocks. The histogram of each of these 3366

blocks is computed, and correlated to the template histogram. It is empirically observed that Most blocks in the image of scanned document have an histogram with a near zero correlation with the template histogram, while the blocks of the CDP have a significantly positive correlation with the template histogram. The block with the highest correlation with the template histogram can generally be assumed to belong to the CDP (especially if the neighboring blocks also exhibit a high correlation). Then, a local search algorithm can be applied to detect all the neighboring blocks that belong to the CDP, and the area containing the CDP can then be cropped and given as input to the restoration function.

ADDITION & Variations

This method may not work to locate copy/counterfeits of lower quality, because the CDP properties can be highly perturbated. In that case, another approach would be to take advantage of the fact that the CDP has generally much more dynamics than any other parts of an image, even if it is copied. To measure the "dynamics" of an area (of eg 50x50 pixels), one could measure the average difference between a pixel and each pixel in the neighborhood.

For any method used, once a block is identified as being part of the CDP, it is still necessary to make a search around that block for all other adjacent blocks that also belong to the CDP. Any local search algorithm can be used to find a set of connected blocks with a given property, where the property is that the block has a "significant" output to the function described above.

4. Distributed CDP (the pixels are grouped into small units which are discontinuous, and are distributed over a large area)

For certain documents, the visual aspect of the CDP can be incompatible with the aesthetics requirements. For example, the aesthetic aspect of a banknote is very important, and in general the security features of these documents must either be unnoticeable, or not disturbing, or naturally fitting with the outlook of the document.

5. More generally, any kind of document (check, package, ID card, etc.) can have aesthetic requirements such that even a small sized CDP might be judged visually unacceptable. It is however possible to split the pixel values of the CDP into small units which are discontinuous, and are distributed over a large area. After scanning the whole document and determining the scale and orientation from one or more detected CDP units, each of the other CDP units can be located, and contribute to the measure of the CDP quality that is compared to the threshold.

Figure 5 below shows a distributed CDP on a check, in which the CDP units are of size 10x10 pixels and regularly distributed every 100 pixels. The number of pixels is equivalent to a CDP of size 240x100 pixels. For this case, the CDP units are very visible, but the goal of this Figure is to show the general idea of distributing the CDP. For more textured CDP, by using smaller CDP units and randomizing the CDP unit location, the aesthetic of the resulting documents can be much less significantly altered.

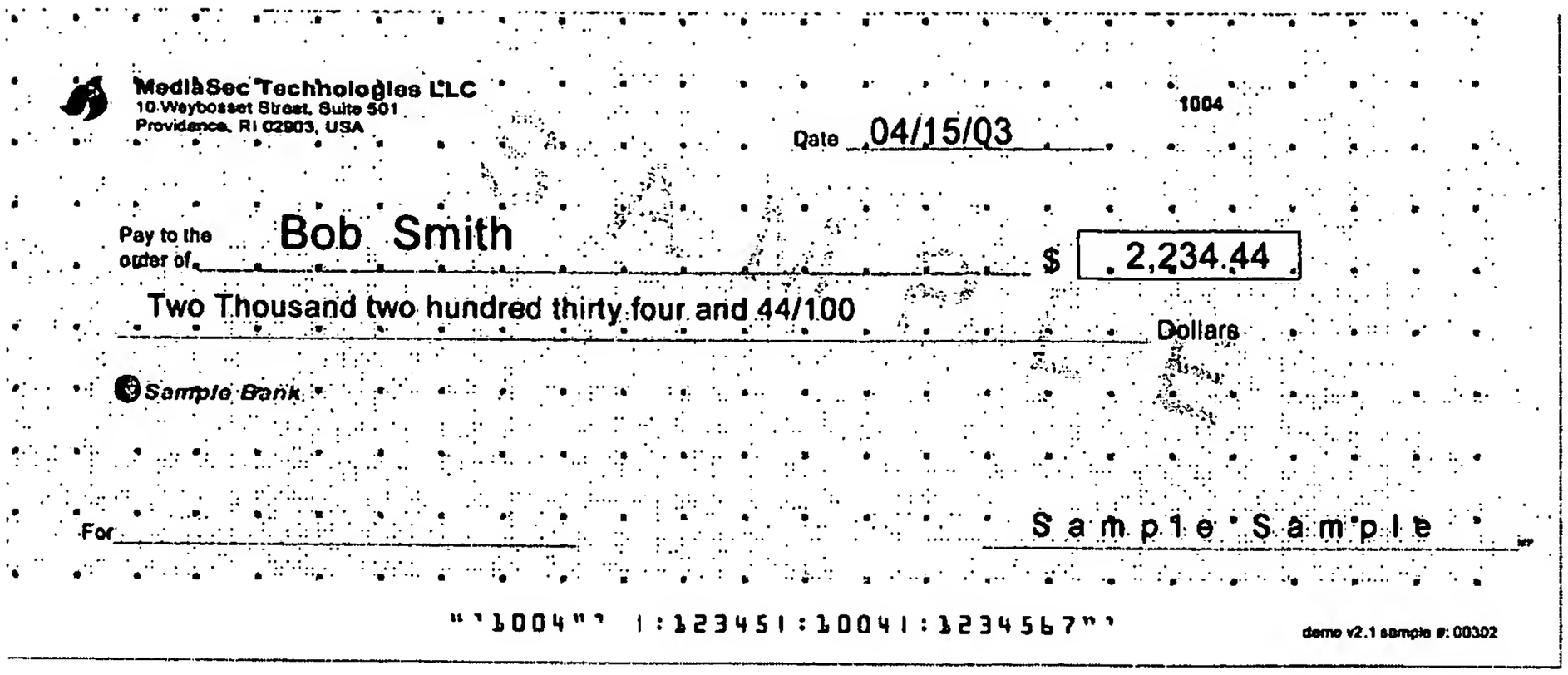


Figure 5: CDP distributed into small units on a check

6. Black/white CDP

Certain type of printers can only print purely black pixels, and do not have the ability to print grayscale CDPs having any possible pixel luminance value between 0 and 255. For these kind of printers, the pseudo-random number generator that is used to generate the CDP pixel values can be set such that a pixel value is either '0' (black), or '1' (white). The algorithms discussed above – points 1,2,3,4- can also be used in the same way for a binary CDP.

It should be remarked that certain printers like inkjet or laser printers actually produce a range of gray tones by employing digital halftoning methods, ie by printing tiny binary dots at a high resolution (eg 1200dpi) for a grayscale image of a lower resolution (eg 300ppi). For those printers, a binary representation of the image is produced by the printer from the input grayscale representation, and in the end it is a binary black/white image that is printed. Instead of generating and printing a grayscale CDP, which has to go through the transformation by the printer, it is possible to produce a higher resolution (eg 1200dpi) binary CDP which is printed as-is by the printer.

Finally, there are other binary printing processes that could print binary CDPs: eg laser engraving, certain holograms, etc.

7. Automate Calibration (training phase).

To optimize the detection of a CDP for a specific application which is defined by fixed settings for the printer/scanner, it is required to evaluate various parameters of the scans of the printouts for that application. For example, the histogram conversion function discussed above depends on the printer and scanner settings, and the quality or similarity threshold used by the detector to take a decision also depends on the scanner settings. However, as the print-scan process is intrinsically noisy and varies –

within statistical boundaries- from print to print and from scan to scan, it is required to estimate statistically the values of the parameters. However, it can be uneasy, tedious and error-prone to print and scan thirty or more CDPs in order to estimate the parameters. On the other side, there are clear advantages to automate this process – required for setting up each application- in order to allow arbitrary people to create applications that detect copies using the CDP.

One way to automate this process is to create a so-called calibration image, a digital image containing several times the same CDP, or CDPs varying by key or payload where the variation is known. One such image is shown in Figure 6. This image is then printed and scanned with the printer and scanner settings of the application. Then the printed calibration image is scanned and process with the CDP calibration software. If the calibration is successful, the calibration software outputs the parameters of the application and the decision threshold of the detector. The decision threshold is typically computed by first measuring the average and variance of the quality index, then by fitting a statistical model to these data in order to determine a threshold that does not result in, for example, 1 error in 10000 cases (assuming the statistical distribution holds). The user can give as a parameter an upper bound on the probability of false alarm (detecting an original as a copy) which will be used in the computation to determine the threshold.

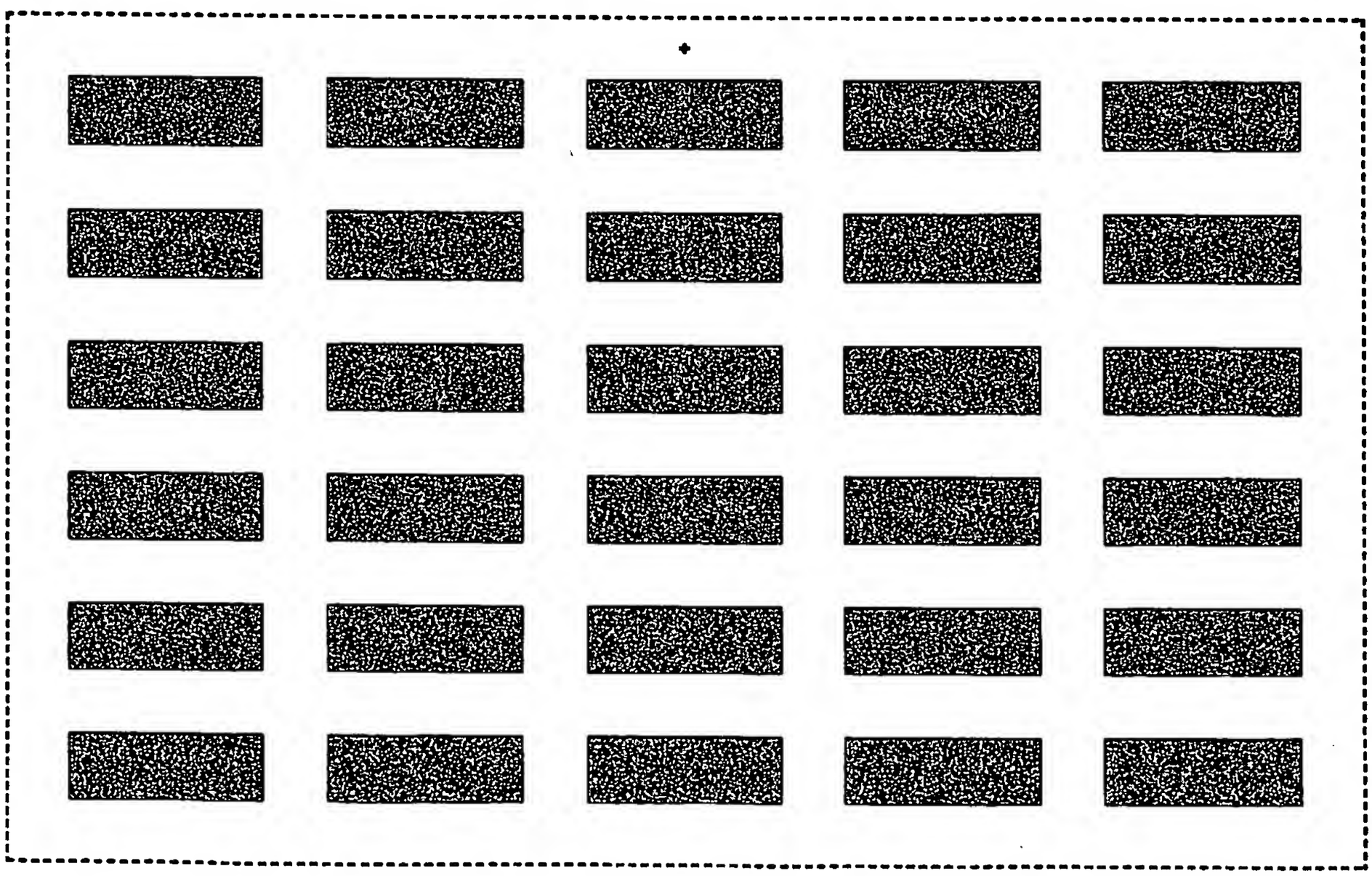


Figure 6: example calibration image

8. Application of the CDP principle to ther analogue media, such as radi frequency, magnet stripe, audio and video.

11

As a recall, the CDP is based on the following two principles:

1) A copy of an original print always goes to through two print-scan process, while an original print only goes to one. There is then a difference of quality between an original print and a copy.

2) The CDP is a key-generated graphic that is designed to be maximally sensitive to print-scanning, and such that an automatic detector working on a digitized representation on the printed CDP can measure an index of quality degradation of the graphic.

It can be easily seen that these two principles can be applied to any analogue media, of which a digital representation exists and was used to produce the analogue version. In all cases, a copy of an analogue media would be made by a digital-analog transformation of the analogue media. And any analogue-media can potentially contain an equivalent of a CDP image, which is equivalent to pseudo-random noisy signal. Let us use the word Copy Detection Signal (CDS) to denote such a copy sensitive signal.

Radio-frequency could keep certain frequency bands to hold the CDS, a certain area of a magnetic stripe could also be kept contain a CDS, and similarly for a video and audio signal could also contain certain area .

9. Distribution of the CDP pattern over a full-page print area while the pattern is very light and serves as a background to the print content.

For letter-size documents, it can be desirable to have a less visible security feature than any CDP and/or a security feature that is distributed everywhere. In point 5 above, it was proposed to divide the CDP into smaller units. It is possible to go a step further by separating the CDP in pixel or dot-size units.

The idea is that dots are spread on the document, and that each dot has a random pixel luminance value. The dots can be regularly spread out or can have random or pseudo-random distribution. The pattern of pseudo-random dots and dot values can be of a fixed size, e.g. 2x2 inches, and tiled over the entire document. An example is given in Figure 7.

At detection time, a dot detection algorithm can be applied to detect all or most of the dots in a digitized image of a printed document. The method described in patent application "Watermarked detection from degraded content" can for example be used. Then, on the set of detected dots, a search algorithm is applied to detect groups of dots with specific properties, that allow to make the registrarttion of the dot patterns and compare them to a digital template. Those properties can be defined by their value, relative position, etc. If detected, these groups of dots allow to make the registration of the dots (i.e. calculate and invert the rotation scaling and translation). Detected dots are matched with the corresponding dots of the digital pattern, and a quality index can be measured.

Obviously, the multiple CDP digital image has to be printed on the printer that will be used in the application, with the same printing parameters, and has to be scanned with the scanner of the application, with the same parameters as well. Not observing this, or using the wrong threshold and parameters, may, may lead to a less reliable detection, if not to completely wrong results.

Figure 7: document image with pseudo-random 2x2 dots containing themselves pseudo-random pixel values.

10. Evaluating the quality of the CDP by using the histogram only, enabling verification without synchronization, since the histogram is RST robust.

In some cases, it might be desirable to make only a rough estimation of whether the document is an original or copy, or it is simply not possible to make a precise estimation: reasons can be too costly computation, corrupted pattern, key or other essential parameters is lacking, or detection software is not made available for security reasons, etc. This is possible by measuring the global properties of the CDP: some global properties are the histogram distribution, the average luminance, the average degree of variation between two consecutive pixels, etc. For measuring these global properties, the step of restoration is not required global properties are generally invariant in rotation, scale and translation-, and it is also not required to have the original digital CDP or the key to generate it. Though the degree of reliability in discerning copies from originals will not be as high if the decision is only based on measuring global properties, most lower quality copies –photocopies, scanning and reprinting with lower end digital imaging equipment- would be detected with such an approach.

Such a detector can also be used to screen suspect documents and take them for further investigation using a detector that makes a more complete analysis of the CDP.

11. Distribution of the CDP pattern into text/graphics area.

The pixels of CDP can be distributed into the text areas in a label or document to make CDP less noticeable. The pixels of CDP form a text or graphics as shown below.

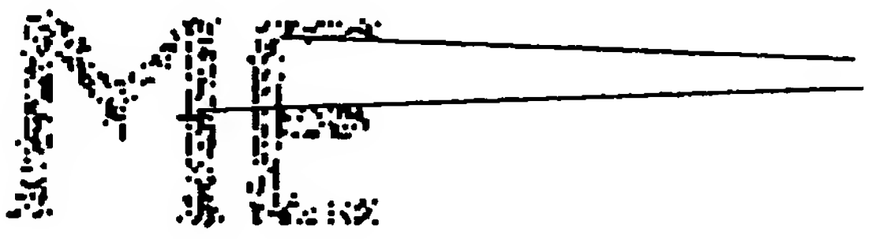


Figure 8. CDP pixels form text or graphics

To verify CDP, just collect the pixels according to the order and measure the quality to make decision.

Fields of Use:

1. **Detection of counterfeit RFID signals (maybe useful in fifteen years!)**

Radio Frequency Identification Devices are attached to individual, arbitrary items, and emit a unique signal that serves as a unique identification of the item. Though the technology is just at its beginning and the cost of RFIDs is still high, various applications of this technology are expected to appear, in particular for applications related to retailing, distribution and storage.

For counterfeiters, the application of RFIDs could mean that counterfeited items must hav counterfeited RFIDs, which emit a counterfeited signal. Counterfeiting a RFID signal could be possible by capturing an original RFID signal, and creating a RFID device that reproduces that captured RFID signal.

The similarity with CDPs to detect counterfeits of printed documents is striking: assuming the RFID signal has the equivalent of an analog form before being emitted, an original signal is captured after one digital-analog-digital conversion, and a counterfeited RFID signal would be captured only after a digital-analog-digital-analog-digital conversion. This additional analog-digital conversion would generally result in an additional loss of quality or information of the RFID signal. It is possible to create a Copy Detection Signal (CDS) that is maximally sensitive to the digital-analog conversions, and that would be emitted by the RFID. This signal would be analyzed the RFID detector to detect if the RFID device authentic or counterfeited.

2. **Used with black/white printers (such as b/w thermo printer)**

See point 6 of algorithm.

3. **Application where an arbitrary or registered individual scans the CDP and sends it to a remote server for verification (remote authentication has been covered in the original application)**

4. **Extend CDP as a measurement for D/A and D/A conversion for all kinds of signals including (wireless) communications, video broadcasting, DVD, ...**

Each A/D and D/A conversion devices will introduce "noise" and/or specific patterns of "noise". By measuring the noises and comparing with the original signal, one can make decision to prevent priracy.

5. **Fill "analog hole" by measuring the quality loss during D/A and A/D conversi n for "Copy Once" of DVD C py Pr tection scheme**

As mentioned above, CDP can also be applied to various types of signals where a digital, fixed form of the signal exists, where the signal is corrupted by transmission (in which it is in "analog" form"), and where it must be re-digitized by the detector. The signal that is stored on a DVD fits the corresponding description. A certain portion of the signal, the "Copy Detection Signal", could be generated in pseudo-random way using a key, it would then be unpredictable and would contain no error correction scheme. This signal would be reproduced, in a degraded way, when a copy of an original DVD is made. If a copy of a copy is made, the signal would be degraded twice, and be of lower quality. It is then possible to attach a device that automatically reads the quality of a DVD when the user wants to make a copy of it, and the copy is made only of the quality of the DVD is judged matching the quality of an original (ie not copied) DVD.

6. Using the payload to be embedded in the CDP to store information about the printer used, enabling dynamic adjustment of the detection device to different printing qualities.

The CDP detector requires parameters that are computed and optimized for a given printer/scanner pair with fixed parameters, also called printer/scanner settings. Two different printer/scanner settings will generally lead to two different values for the parameters, which may be incompatible. Though for many applications, the printer-scanner settings can be assumed to be fixed and known, there are some applications where it cannot be known in advance which printer and/or scanner settings will be used at detection time. For example, in applications where the CDP can be printed from virtually any printer, it may not be known by the detector which set of parameters should be used for detection. It is however possible to store in the CDP the parameters that must be used by the detector (see eg point 2 of algorithms). For example, the quality threshold (over/under which a CDP can be judged as original/copy) or the histogram conversion function can be stored in the CDP.

7. Universal document authentication (authentic always): every printer will print CDP onto each document being printed.

A CDP can be printed per default on each document being printed on any of a set of printers, enabling to potentially detect any copy of the document.

8. Lock CDP with the security features

A physical security feature (such as a fiber, inks) is first detected and then used as a key for CDP.

CDP can also be combined with security storage devices such as smart cards, 2D barcode, magnetic card, ... The secret key and calibration data (which depends on the printer properties) can be stored in those storage devices.

CDP is also complementary with physical copy prevention features such as "VOID" appearing once copied.

9. Support multiple printers and scanners

Like a printer/scanner driver, a CDP reader (verifier) can select appropriate calibration data manually or automatically. For example, when a CDP is printed, a code which uniquely identify the printer or the category (printer model) of the printer is embedded into the CDP or printed/stored into a database or on the document where CDP is printed. At the verification stage, the reader first detect the code and then select the appropriate calibration data for verification.

The scanner-dependent parameters such as thresholds for verification can be automatically selected in the similar way.

10. Postage meter applications

A CDP can be automatically inserted in the digital image of the postal indicia, which is then typically printed on a sticker glued to the envelope, or directly printed on the envelope. That CDP can later be used to automatically detect if a postal indicia is an original or a copy, e.g. made by scanning and reprinting an original indicia. The quality index measured on the CDP can be combined with other features of the postal indicia (determining the printer that produced it, analyzing the font of the letter, reading a printed digital watermark) into a global score that is used to automatically or manually determine if the document is a copy.

The CDP can be used for forensic verification, e.g. when a suspicious postal indicia is brought to a station equipped by a flatbed scanner and the detection software. It can also be used for automatic verification when a high-speed scanner (e.g. a WFOV) makes an image capture of each postal indicia.

In such an application, the key for the CDP can be fixed, or variable. If it is variable, the key can be (partly) derived from other information contained in the postal indicia, e.g. the sender name.

11. Valuable documents (bills, tickets and certificates, etc.)

Bills, event tickets and transportation tickets and other valuable documents can often be copied with high-quality copier or scanner-printer.

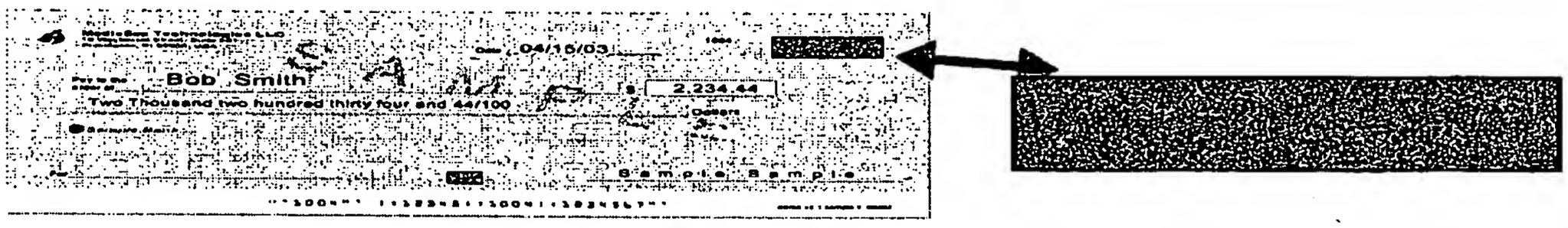
12. Check fraud application

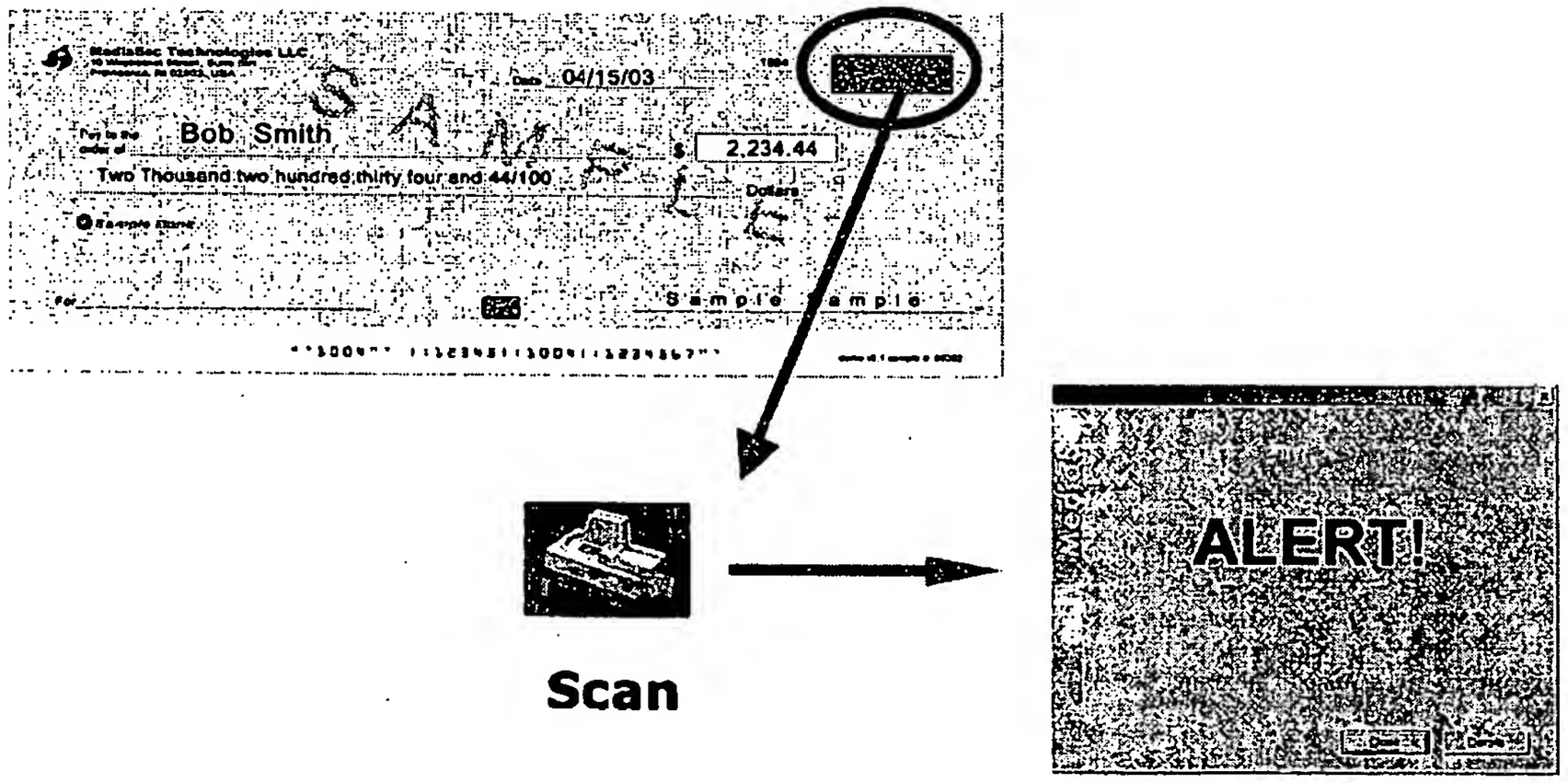
If fraudulent duplication (photocopying or scanning-printing) of checks is a concern, a secure pattern called a Copy Detection Pattern (CDP) can be printed on check stock. The CDP is incorporated into the design of the check.

Simple in appearance, a CDP is actually an extremely complex digital image.

The verification process measures even the slightest amount of degradation occurring during the copying process.

From this measured degradation, copies can be discerned from originals with extreme reliability.





Provisional Patent Application

Patent Application for

Visible authentication patterns for printed documents

Extension II

Inventors: Justin Picard, Dr. Jian Zhao

Assignee: MediaSec Technologies GmbH

Visible authentication patterns for printed documents

Cross references to related applications

The present application is the second extension and improvement of PCT/US03/15188, Visible authentication patterns for printed documents, filed 14 May 2003.

Extensions of Algorithms:

1. Histogram conversion as a core part of the algorithm

As already discussed in previous extension, for certain applications, certain variations can occur in the printing-scanning environment. This variations can typically occur in:

- different printers used to print the document containing the CDP
- different substrates or papers used to apply the CDP or overlaid on the CDP
- different scanners used to scan the CDP
- different treatments to the document or object that bears the CDP

One such application would be a postage meter where CDPs can be printed on different types of envelopes (made of paper with different properties) and sometimes from different printers.

For such applications, the histogram conversion function could be different for each combination of the properties mentioned above. Applying the wrong histogram conversion function would result generally in a distortion of the distance between the test and digital CDP. One solution **suggested in the previous patent application** would then be to use several histogram conversion functions, and when verifying a CDP, try each of the histogram conversion function and consider only the CDP with the lowest distance to the digital CDP. If other parameters are dependent on the specific printer-paper-scanner combination (e.g. the threshold), then the selection of the "best" histogram conversion function would imply the selection of the other parameters.

There are other solutions to minimize the impact of an unstable printing-scanning environment that can be used. Considering the set of possible histogram conversion functions for one application, where one conversion function can be generated for each printed-scanned CDP, there are different ways to integrate the natural variation. One solution consists in measuring the standard deviation for each pixel, and then use it as a normalizing factor considered when measuring a distance between the scanned-converted CDP and the digital CDP. Another one consists in estimating an upper and lower bound for the typical value of each pixel luminance (e.g. luminance 100 should have a typical value between 90 and 110 in the scanned image, while luminance 40 could have a typical value between 20 and 60 –which is twice as large) penalizing more severely the pixels that are out of those bounds. Yet another solution would consists in having a set of different conversion functions that represent faithfully the spectrum of different conversion functions that can occur in the application, and using the most adequate one – resulting in the smallest distance between the converted and digital CDP - each time.

Still, some more conservative solutions can be used, one being to not use pre-estimated parameters, but to estimate them on the tested CDP. This allows for more tolerance to variations but there are two potential limitations to this approach:

- As parameters are estimated on the current image, this allows more tolerance for copies which do not necessarily have to respect the typical histogram(s) of an original CDP

- as parameters are estimated from only one image, one can expect a less precise estimation of parameters. For example, for a 10000 pixels CDP with 255 equiprobable luminance values, there would be on average less than 40 samples per luminance value, and by natural statistical variation some luminance would have significantly less than 40 samples.

The first problem can be treated in different ways. One is to have a pre-estimated conversion function (as in the standard approach), but to use it not to convert the scanned CDP, but only to measure a distance to the conversion function estimated for the current CDP. One possible measures of distance between conversion functions are:

D(f',f)=1/256 *Sum(lum=0 to lum=255) abs(f'(lum)-f(lum))

Where f'(.) and f(.) are respectively the self-estimated conversion function and the average conversion function. It is also possible to enter a normalization factor g(lum) corresponding to the natural variation for each pixel, e.g.:

Dn(f',f)=1/256 * Sum(lum=0 to lum=255) abs(f'(lum)-f(lum))/g(lum)

This distance can be used as additional evidence that can enter in the decision. For example, two different test CDPs might have the same quality index of 78 when measuring their distance to the digital CDP with the self-estimated parameter conversion. However, their self-estimated conversion function might have a different distance to the average conversion function, e.g. Dn(f',f)=2.5 for the first CDP and D(f',f)=0.5 for the second CDP. This higher distance for the first CDP might be used to derive that it is a copy (despite its high quality), while for the same quality index the second CDP would be considered an original.

The second problem can be treated by assuming a model for the conversion function, i.e. that the conversion function follows a certain regression function, e.g. a polynomial function or a logistic regression. This minimizes the number of parameters to be estimated, and gives smoother functions with no discontinuity.

Finally, in some cases the printing-scanning properties can be evolving in time, and/or it is not feasible to calibrate the CDP detector initially. In those cases, a flexible approach consists in incorporating the data of each new scan, allowing more tolerance in the beginning when the parameters of the environment are not known, and progressively decreasing that tolerance as the addition of new data allows for a more precise estimation of the underlying parameters. This approach is general in nature, and is valid for all parameters relevant to the CDP detection (conversion function, threshold, etc.). To incorporate new knowledge, Bayesian learning can be applied where the importance assigned to the priors is progressively decreased. The information gained by evaluation can be stored in a database and shared between different verification stations. This approach allows for decoupling of the information regarding the CDP quality at the time of printing from the scanning parameters. The information is evaluated as the verification occurs enabling a more flexible integration process of the solution.

A combination of the methods above is possible, in a scenario where several of the evaluation strategies are applied and the result is weighted to derive a probability for the resulting decision of the CDP quality.

20

Information about the print-scan environment and the properties that might affect the CDP quality (see above) can be stored in an encoded, machine readable way on the document. Alternatively it can be encoded in the CDP.

2. Method of detecting the position of a CDP on a document with no a priori knowledge on its location

In the previous extension, a method is suggested for "detecting the position of a CDP on a document with no a priori knowledge on its location". It is proposed to use the typical histogram of the CDP for that purpose, which is assumed to be a unique statistical descriptor of areas containing the CDP. However, the implementation of that approaches required having some a priori knowledge on the typical CDP histogram for the given printing-scanning environment; it might not always be possible to have such knowledge. In that case, it is possible to use a property general to all CDPs, that their pixel values of the maximum entropy in the digital domain (8 bits/pixel for a grayscale image). Though this entropy is decreased by printing-scanning, it is often the case that areas with the CDP still are the highest entropy area in an image. Therefore, by measuring the entropy in each area of the scanned image and selecting the area with highest entropy, it is possible to derive the location of the CDP.

In some cases, this approach does not work because of order highly textured areas in the image in which the texture –and therefore the entropy- have been better preserved during the printing-scanning. In this case, one approach to avoid detecting unwanted areas that have high entropy, consists in setting some restrictions on the set of possible pixel values that enter in the computation of the entropy. For example, if CDP s often have pixel values between 0 and 150 in the scanned image, one can exclude from the computation of the entropy all pixel values that have a luminance higher than 150.

3. Method of using information about the print-scan process to derive the digital CDP used for comparison

Typically, the scanned CDP is compared to the digital CDP that was used at creation time. This digital CDP however does not consider the natural, average effect that occur during printing scanning, which can typically be described by low-pass or band-pass filtering. Comparing the scanned CDP to a digital CDP that simulates the printing and scanning effects may allow for a more precise measurement of the quality index of a given CDP. For example, as a general observation, a pixel with a low luminance in the digital CDP will, of course, generally translate into a pixel that has a low luminance in the scanned CDP; however, its luminance will generally be higher if it is surrounded by bright pixels than by dark pixels. Using a digital CDP that has gone through a simulated print-scan process, this impact of the neighborhood can be taken into account.

There are several ways to estimate the impact of the print-scan process. They can generally be classified into three categories:

1. The print-scan process can be simulated by looking at the average luminance value of each pixel in a large number of scanned CDPs.

2. The print-scan process can be estimated as a filter with a specific frequency response. The filter is then applied to the digital CDP, and the resulting image is used in the detection process

3. The physical properties of the printer, the paper and the scanner can sometimes be known, and can be used to estimate the pixel values of the scanned CDP. For example, thermal printers typically have a residual heat after printing each individual pixel or dot; this residual heat may have an impact on the following pixel, which could be printed with excessive heat resulting in a darker than desired pixel. These physical effects, and many others pertaining to the specific printing and scanning device, and to the ink absorption properties of the paper, can be considered.

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING
SUBMISSION OR TRANSMITTAL
OF PRIORITY DOCUMENT

(PCT Administrative Instructions, Section 411)

To:

NELSON, Gordon, E.
57 Central Street
P.O. Box 782
Rowley, MA 01969
ETATS-UNIS D'AMERIQUE

Date of mailing *(day/month/year)* 24 March 2005 (24.03.2005)	
Applicant's or agent's file reference medias01.022	**IMPORTANT NOTIFICATION**
International application No. PCT/US05/000201	International filing date *(day/month/year)* 06 January 2005 (06.01.2005)
International publication date *(day/month/year)*	Priority date *(day/month/year)* 06 January 2004 (06.01.2004)
Applicant MEDIASEC TECHNOLOGIES, GMBH et al	

1. By means of this Form, which replaces any previously issued notification concerning submission or transmittal of priority documents, the applicant is hereby notified of the date of receipt by the International Bureau of the priority document(s) relating to all earlier application(s) whose priority is claimed. Unless otherwise indicated by the letters "NR", in the right-hand column or by an asterisk appearing next to a date of receipt, the priority document concerned was submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b).

2. *(If applicable)* The **letters "NR"** appearing in the right-hand column denote **a priority document which, on the date of mailing of this Form, had not yet been received by the International Bureau** under Rule 17.1(a) or (b). Where, under Rule 17.1(a), the priority document must be submitted by the applicant to the receiving Office or the International Bureau, but the applicant fails to submit the priority document within the applicable time limit under that Rule, **the attention of the applicant is directed** to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

3. *(If applicable)* **An asterisk (*)** appearing next to a date of receipt, in the right-hand column, denotes **a priority document submitted or transmitted to the International Bureau but not in compliance with Rule 17.1(a) or (b)** (the priority document was received after the time limit prescribed in Rule 17.1(a) or the request to prepare and transmit the priority document was submitted to the receiving Office after the applicable time limit under Rule 17.1(b)). Even though the priority document was not furnished in compliance with Rule 17.1(a) or (b), the International Bureau will nevertheless transmit a copy of the document to the designated Offices, for their consideration. In case such a copy is not accepted by the designated Office as the priority document, Rule 17.1(c) provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

Priority date	Priority application No.	Country or regional Office or PCT receiving Office	Date of receipt of priority document
06 January 2004 (06.01.2004)	60/534,520	US	03 March 2005 (03.03.2005)

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. +41 22 740 14 35

Authorized officer

Giffo Schmitt Beate

Facsimile No. +41 22 338 87 20
Telephone No. +41 22 338 9241